Duality
DM par email: dm.daspremont@gmail.com
Outline

- Lagrange dual problem
- weak and strong duality
- geometric interpretation
- optimality conditions
- perturbation and sensitivity analysis
- examples
- theorems of alternatives
- generalized inequalities
Lagrangian

standard form problem (not necessarily convex)

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0, \quad i = 1, \ldots, p
\end{align*}
\]

variable \( x \in \mathbb{R}^n \), domain \( D \), optimal value \( p^* \)

Lagrangian: \( L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R} \), with \( \text{dom} \ L = D \times \mathbb{R}^m \times \mathbb{R}^p \),

\[
L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)
\]

- weighted sum of objective and constraint functions
- \( \lambda_i \) is Lagrange multiplier associated with \( f_i(x) \leq 0 \)
- \( \nu_i \) is Lagrange multiplier associated with \( h_i(x) = 0 \)
Lagrange dual function

Lagrange dual function: \( g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \),

\[
g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu)
\]

\[
= \inf_{x \in D} \left( f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)
\]

\( g \) is concave, can be \(-\infty\) for some \( \lambda, \nu \)

**lower bound property:** if \( \lambda \succeq 0 \), then \( g(\lambda, \nu) \leq p^* \)

proof: if \( \bar{x} \) is feasible and \( \lambda \succeq 0 \), then

\[
f_0(\bar{x}) \geq L(\bar{x}, \lambda, \nu) \geq \inf_{x \in D} L(x, \lambda, \nu) = g(\lambda, \nu)
\]

minimizing over all feasible \( \bar{x} \) gives \( p^* \geq g(\lambda, \nu) \)
Least-norm solution of linear equations

\[
\begin{align*}
\text{minimize} & \quad x^T x \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

dual function

- Lagrangian is \( L(x, \nu) = x^T x + \nu^T (Ax - b) \)
- to minimize \( L \) over \( x \), set gradient equal to zero:
  \[
  \nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \implies \quad x = -\left(\frac{1}{2}\right)A^T \nu
  \]
- plug in in \( L \) to obtain \( g \):
  \[
  g(\nu) = L\left(-\frac{1}{2}A^T \nu, \nu\right) = -\frac{1}{4} \nu^T AA^T \nu - b^T \nu
  \]
  a concave function of \( \nu \)

lower bound property: \( p^* \geq -\left(\frac{1}{4}\right) \nu^T AA^T \nu - b^T \nu \) for all \( \nu \)
Standard form LP

minimize \( c^T x \)
subject to \( Ax = b, \quad x \succeq 0 \)

dual function

- Lagrangian is

\[
L(x, \lambda, \nu) = c^T x + \nu^T (Ax - b) - \lambda^T x \\
= -b^T \nu + (c + A^T \nu - \lambda)^T x
\]

- \( L \) is linear in \( x \), hence

\[
g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} 
- b^T \nu & A^T \nu - \lambda + c = 0 \\
- \infty & \text{otherwise}
\end{cases}
\]

\( g \) is linear on affine domain \( \{ (\lambda, \nu) \mid A^T \nu - \lambda + c = 0 \} \), hence concave

lower bound property: \( p^* \geq - b^T \nu \) if \( A^T \nu + c \succeq 0 \)
Equality constrained norm minimization

\[
\begin{align*}
\text{minimize} & \quad \|x\| \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

dual function

\[
g(\nu) = \inf_x (\|x\| - \nu^T Ax + b^T \nu) = \begin{cases} 
\nu^T b & \|A^T \nu\|_* \leq 1 \\
-\infty & \text{otherwise}
\end{cases}
\]

where \(\|\nu\|_* = \sup_{\|u\| \leq 1} u^T \nu\) is dual norm of \(\|\cdot\|\)

proof: follows from \(\inf_x (\|x\| - y^T x) = 0\) if \(\|y\|_* \leq 1\), \(-\infty\) otherwise

- if \(\|y\|_* \leq 1\), then \(\|x\| - y^T x \geq 0\) for all \(x\), with equality if \(x = 0\)
- if \(\|y\|_* > 1\), choose \(x = tu\) where \(\|u\| \leq 1\), \(u^T y = \|y\|_* > 1\):

\[
\|x\| - y^T x = t(\|u\| - \|y\|_*) \to -\infty \quad \text{as} \ t \to \infty
\]

lower bound property: \(p^* \geq b^T \nu\) if \(\|A^T \nu\|_* \leq 1\)
Two-way partitioning

\[
\begin{align*}
\text{minimize} & \quad x^T W x \\
\text{subject to} & \quad x_i^2 = 1, \quad i = 1, \ldots, n
\end{align*}
\]

- a nonconvex problem; feasible set contains \(2^n\) discrete points
- interpretation: partition \(\{1, \ldots, n\}\) in two sets; \(W_{ij}\) is cost of assigning \(i, j\) to the same set; \(-W_{ij}\) is cost of assigning to different sets

**dual function**

\[
g(v) = \inf_{x} (x^T W x + \sum_i v_i(x_i^2 - 1)) = \inf_{x} x^T (W + \text{diag}(v)) x - 1^T v
\]

\[
= \begin{cases} 
-1^T v & W + \text{diag}(v) \succeq 0 \\
-\infty & \text{otherwise}
\end{cases}
\]

**lower bound property:** \(p^* \geq -1^T v\) if \(W + \text{diag}(v) \succeq 0\)

example: \(v = -\lambda_{\min}(W)1\) gives bound \(p^* \geq n\lambda_{\min}(W)\)
The dual problem

Lagrange dual problem

\[
\begin{align*}
\text{maximize} & \quad g(\lambda, \nu) \\
\text{subject to} & \quad \lambda \succeq 0
\end{align*}
\]

- finds best lower bound on \( p^* \), obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted \( d^* \)
- \( \lambda, \nu \) are dual feasible if \( \lambda \succeq 0, (\lambda, \nu) \in \text{dom} \ g \)
- often simplified by making implicit constraint \( (\lambda, \nu) \in \text{dom} \ g \) explicit

example: standard form LP and its dual (page 8)

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b \\
x & \succeq 0
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad -b^T \nu \\
\text{subject to} & \quad A^T \nu + c \succeq 0
\end{align*}
\]
Weak and strong duality

**weak duality:** $d^* \leq p^*$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems
  for example, solving the SDP

\[
\begin{align*}
\text{maximize} & \quad -1^T \nu \\
\text{subject to} & \quad W + \text{diag}(\nu) \succeq 0
\end{align*}
\]

  gives a lower bound for the two-way partitioning problem on page 10

**strong duality:** $d^* = p^*$

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**
Slater’s constraint qualification

strong duality holds for a convex problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

if it is strictly feasible, i.e.,

\[
\exists x \in \text{int} \mathcal{D} : \quad f_i(x) < 0, \quad i = 1, \ldots, m, \quad Ax = b
\]

- also guarantees that the dual optimum is attained (if \( p^* > -\infty \))
- can be sharpened: e.g., can replace \( \text{int} \mathcal{D} \) with \( \text{relint} \mathcal{D} \) (interior relative to affine hull); linear inequalities do not need to hold with strict inequality, . . .
- there exist many other types of constraint qualifications
Feasibility problems

**feasibility problem A** in $x \in \mathbb{R}^n$.

\[
 f_i(x) < 0, \quad i = 1, \ldots, m, \quad h_i(x) = 0, \quad i = 1, \ldots, p
\]

**feasibility problem B** in $\lambda \in \mathbb{R}^m$, $\nu \in \mathbb{R}^p$.

\[
 \lambda \succeq 0, \quad \lambda \neq 0, \quad g(\lambda, \nu) \geq 0
\]

where $g(\lambda, \nu) = \inf_x (\sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x))$

- feasibility problem B is convex ($g$ is concave), even if problem A is not
- A and B are always **weak alternatives**: at most one is feasible
  proof: assume $\bar{x}$ satisfies A, $\lambda$, $\nu$ satisfy B
  \[
  0 \leq g(\lambda, \nu) \leq \sum_{i=1}^m \lambda_i f_i(\bar{x}) + \sum_{i=1}^p \nu_i h_i(\bar{x}) < 0
  \]
- A and B are **strong alternatives** if exactly one of the two is feasible (can prove infeasibility of A by producing solution of B and vice-versa).
**Inequality form LP**

**primal problem**

\[
\begin{align*}
& \text{minimize} \quad c^T x \\
& \text{subject to} \quad Ax \preceq b
\end{align*}
\]

**dual function**

\[
g(\lambda) = \inf_x \left( (c + A^T \lambda)^T x - b^T \lambda \right) = \begin{cases} 
- b^T \lambda & A^T \lambda + c = 0 \\
- \infty & \text{otherwise}
\end{cases}
\]

**dual problem**

\[
\begin{align*}
& \text{maximize} \quad -b^T \lambda \\
& \text{subject to} \quad A^T \lambda + c = 0, \quad \lambda \succeq 0
\end{align*}
\]

- from Slater's condition: \( p^* = d^* \) if \( A\tilde{x} \prec b \) for some \( \tilde{x} \)
- in fact, \( p^* = d^* \) except when primal and dual are infeasible
Quadratic program

**primal problem** (assume $P \in S^{n}_{++}$)

\[
\begin{align*}
\text{minimize} & \quad x^T P x \\
\text{subject to} & \quad A x \preceq b
\end{align*}
\]

**dual function**

\[
g(\lambda) = \inf_x (x^T P x + \lambda^T (A x - b)) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda
\]

**dual problem**

\[
\begin{align*}
\text{maximize} & \quad -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda \\
\text{subject to} & \quad \lambda \succeq 0
\end{align*}
\]

- from Slater’s condition: $p^* = d^*$ if $A \tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^* = d^*$ always
A nonconvex problem with strong duality

\[
\begin{align*}
\text{minimize} & \quad x^T A x + 2 b^T x \\
\text{subject to} & \quad x^T x \leq 1
\end{align*}
\]

nonconvex if \( A \not\preceq 0 \)

dual function: \( g(\lambda) = \inf_x (x^T (A + \lambda I)x + 2 b^T x - \lambda) \)

- unbounded below if \( A + \lambda I \not\succeq 0 \) or if \( A + \lambda I \succeq 0 \) and \( b \not\in \mathcal{R}(A + \lambda I) \)
- minimized by \( x = -(A + \lambda I)^\dagger b \) otherwise: \( g(\lambda) = -b^T (A + \lambda I)^\dagger b - \lambda \)

dual problem and equivalent SDP:

\[
\begin{align*}
\text{maximize} & \quad -b^T (A + \lambda I)^\dagger b - \lambda \\
\text{subject to} & \quad A + \lambda I \succeq 0 \\
& \quad b \in \mathcal{R}(A + \lambda I)
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad -t - \lambda \\
\text{subject to} & \quad \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0
\end{align*}
\]

strong duality although primal problem is not convex (not easy to show)
Geometric interpretation

For simplicity, consider problem with one constraint $f_1(x) \leq 0$

interpretation of dual function:

$$g(\lambda) = \inf_{(u,t) \in \mathcal{G}} (t + \lambda u), \quad \text{where} \quad \mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in D\}$$

- $\lambda u + t = g(\lambda)$ is (non-vertical) supporting hyperplane to $\mathcal{G}$
- hyperplane intersects $t$-axis at $t = g(\lambda)$
**epigraph variation:** same interpretation if \( G \) is replaced with

\[
\mathcal{A} = \{(u, t) \mid f_1(x) \leq u, f_0(x) \leq t \text{ for some } x \in \mathcal{D}\}
\]

**strong duality**

- holds if there is a non-vertical supporting hyperplane to \( \mathcal{A} \) at \((0, p^*)\)
- for convex problem, \( \mathcal{A} \) is convex, hence has supp. hyperplane at \((0, p^*)\)
- Slater’s condition: if there exist \((\bar{u}, \bar{t}) \in \mathcal{A} \) with \(\bar{u} < 0\), then supporting hyperplanes at \((0, p^*)\) must be non-vertical
Complementary slackness

Assume strong duality holds, $x^*$ is primal optimal, $(\lambda^*, \nu^*)$ is dual optimal

\[
f_0(x^*) = g(\lambda^*, \nu^*) = \inf_x \left( f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right)
\]

\[
\leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*)
\]

\[
\leq f_0(x^*)
\]

hence, the two inequalities hold with equality

- $x^*$ minimizes $L(x, \lambda^*, \nu^*)$
- $\lambda_i^* f_i(x^*) = 0$ for $i = 1, \ldots, m$ (known as complementary slackness):

\[
\lambda_i^* > 0 \implies f_i(x^*) = 0, \quad f_i(x^*) < 0 \implies \lambda_i^* = 0
\]
Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable $f_i, h_i$):

1. **Primal feasibility**: $f_i(x) \leq 0, \ i = 1, \ldots, m, \ h_i(x) = 0, \ i = 1, \ldots, p$

2. **Dual feasibility**: $\lambda \succeq 0$

3. **Complementary slackness**: $\lambda_i f_i(x) = 0, \ i = 1, \ldots, m$

4. Gradient of Lagrangian with respect to $x$ vanishes (**first order condition**):

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

If strong duality holds and $x, \lambda, \nu$ are optimal, then they must satisfy the KKT conditions
KKT conditions for convex problem

If $\bar{x}$, $\bar{\lambda}$, $\bar{\nu}$ satisfy KKT for a **convex problem**, then they are optimal:

- from complementary slackness: $f_0(\bar{x}) = L(\bar{x}, \bar{\lambda}, \bar{\nu})$
- from 4th condition (and convexity): $g(\bar{\lambda}, \bar{\nu}) = L(\bar{x}, \bar{\lambda}, \bar{\nu})$

hence, $f_0(\bar{x}) = g(\bar{\lambda}, \bar{\nu})$

If **Slater’s condition** is satisfied, $x$ is optimal if and only if there exist $\lambda$, $\nu$ that satisfy KKT conditions

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem

Summary:

- When strong duality holds, the KKT conditions are necessary conditions for optimality
- If the problem is **convex**, they are also sufficient
example: water-filling (assume $\alpha_i > 0$)

minimize $-\sum_{i=1}^{n} \log(x_i + \alpha_i)$
subject to $x \succeq 0$, $1^T x = 1$

$x$ is optimal iff $x \succeq 0$, $1^T x = 1$, and there exist $\lambda \in \mathbb{R}^n$, $\nu \in \mathbb{R}$ such that

\[\lambda \succeq 0, \quad \lambda_i x_i = 0, \quad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu\]

- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$
- determine $\nu$ from $1^T x = \sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\} = 1$

interpretation

- $n$ patches; level of patch $i$ is at height $\alpha_i$
- flood area with unit amount of water
- resulting level is $1/\nu^*$
Perturbation and sensitivity analysis

(uncerturbed) optimization problem and its dual

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0, \quad i = 1, \ldots, p
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad g(\lambda, \nu) \\
\text{subject to} & \quad \lambda \succeq 0
\end{align*}
\]

perturbed problem and its dual

\[
\begin{align*}
\text{min} & \quad f_0(x) \\
\text{s.t.} & \quad f_i(x) \leq u_i, \quad i = 1, \ldots, m \\
& \quad h_i(x) = v_i, \quad i = 1, \ldots, p
\end{align*}
\]

\[
\begin{align*}
\text{max} & \quad g(\lambda, \nu) - u^T \lambda - v^T \nu \\
\text{s.t.} & \quad \lambda \succeq 0
\end{align*}
\]

- $x$ is primal variable; $u$, $v$ are parameters
- $p^*(u, v)$ is optimal value as a function of $u$, $v$
- we are interested in information about $p^*(u, v)$ that we can obtain from the solution of the unperturbed problem and its dual
Global sensitivity result: Strong duality holds for unperturbed problem and $\lambda^*$, $\nu^*$ are dual optimal for unperturbed problem. Apply weak duality to perturbed problem:

$$p^*(u, v) \geq g(\lambda^*, \nu^*) - u^T\lambda^* - v^T\nu^*$$

$$= p^*(0, 0) - u^T\lambda^* - v^T\nu^*$$

Local sensitivity: if (in addition) $p^*(u, v)$ is differentiable at $(0, 0)$, then

$$\lambda^*_i = -\frac{\partial p^*(0, 0)}{\partial u_i}, \quad \nu^*_i = -\frac{\partial p^*(0, 0)}{\partial v_i}$$
Duality and problem reformulations

- equivalent formulations of a problem can lead to very different duals
- reformulating the primal problem can be useful when the dual is difficult to derive, or uninteresting

Common reformulations

- introduce new variables and equality constraints
- make explicit constraints implicit or vice-versa
- transform objective or constraint functions
  - *e.g.*, replace $f_0(x)$ by $\phi(f_0(x))$ with $\phi$ convex, increasing
Introducing new variables and equality constraints

minimize \( f_0(Ax + b) \)

- dual function is constant: \( g = \inf_x L(x) = \inf_x f_0(Ax + b) = p^* \)
- we have strong duality, but dual is quite useless

**reformulated problem and its dual**

minimize \( f_0(y) \) subject to \( Ax + b - y = 0 \)

maximize \( b^T \nu - f_0^*(\nu) \) subject to \( A^T \nu = 0 \)

dual function follows from

\[
g(\nu) = \inf_{x,y} (f_0(y) - \nu^T y + \nu^T Ax + b^T \nu) \\
= \begin{cases} 
-f_0^*(\nu) + b^T \nu & A^T \nu = 0 \\
-\infty & \text{otherwise}
\end{cases}
\]
norm approximation problem: minimize $\|Ax - b\|$

minimize $\|y\|
subject to $y = Ax - b$

can look up conjugate of $\| \cdot \|$, or derive dual directly

$$g(\nu) = \inf_{x,y} (\|y\| + \nu^T y - \nu^T Ax + b^T \nu)$$

$$= \begin{cases} 
  b^T \nu + \inf_y (\|y\| + \nu^T y) & A^T \nu = 0 \\
  -\infty & \text{otherwise}
\end{cases}$$

$$= \begin{cases} 
  b^T \nu & A^T \nu = 0, \quad \|\nu\|_* \leq 1 \\
  -\infty & \text{otherwise}
\end{cases}$$

(see page 7)

dual of norm approximation problem

maximize $b^T \nu$
subject to $A^T \nu = 0, \quad \|\nu\|_* \leq 1$
Implicit constraints

**LP with box constraints:** primal and dual problem

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b \\
& \quad -1 \leq x \leq 1
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad -b^T \nu - 1^T \lambda_1 - 1^T \lambda_2 \\
\text{subject to} & \quad c + A^T \nu + \lambda_1 - \lambda_2 = 0 \\
& \quad \lambda_1 \geq 0, \quad \lambda_2 \geq 0
\end{align*}
\]

reformulation with box constraints made implicit

\[
\begin{align*}
\text{minimize} & \quad f_0(x) = \begin{cases} 
  c^T x & -1 \leq x \leq 1 \\
  \infty & \text{otherwise}
\end{cases} \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

dual function

\[
g(\nu) = \inf_{-1 \leq x \leq 1} (c^T x + \nu^T (Ax - b)) = -b^T \nu - \|A^T \nu + c\|_1
\]

dual problem: maximize \(-b^T \nu - \|A^T \nu + c\|_1\)
Problems with generalized inequalities

\[
\begin{align*}
& \text{minimize} \quad f_0(x) \\
& \text{subject to} \quad f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0, \quad i = 1, \ldots, p
\end{align*}
\]

\[\preceq_{K_i}\] is generalized inequality on \(\mathbb{R}^{k_i}\)

definitions are parallel to scalar case:

- Lagrange multiplier for \(f_i(x) \preceq_{K_i} 0\) is vector \(\lambda_i \in \mathbb{R}^{k_i}\)
- Lagrangian \(L : \mathbb{R}^n \times \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \to \mathbb{R}\), is defined as

\[
L(x, \lambda_1, \ldots, \lambda_m, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i^T f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)
\]

- dual function \(g : \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \to \mathbb{R}\), is defined as

\[
g(\lambda_1, \ldots, \lambda_m, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda_1, \ldots, \lambda_m, \nu)
\]
lower bound property: if \( \lambda_i \succeq K_i^* 0 \), then \( g(\lambda_1, \ldots, \lambda_m, \nu) \leq p^* \)

proof: if \( \bar{x} \) is feasible and \( \lambda \succeq K_i^* 0 \), then

\[
\begin{align*}
    f_0(\bar{x}) &\geq f_0(\bar{x}) + \sum_{i=1}^{m} \lambda_i^T f_i(\bar{x}) + \sum_{i=1}^{p} \nu_i h_i(\bar{x}) \\
    &\geq \inf_{x \in \mathcal{D}} L(x, \lambda_1, \ldots, \lambda_m, \nu) \\
    &= g(\lambda_1, \ldots, \lambda_m, \nu)
\end{align*}
\]

minimizing over all feasible \( \bar{x} \) gives \( p^* \geq g(\lambda_1, \ldots, \lambda_m, \nu) \)

dual problem

\[
\begin{align*}
\text{maximize} & \quad g(\lambda_1, \ldots, \lambda_m, \nu) \\
\text{subject to} & \quad \lambda_i \succeq K_i^* 0, \quad i = 1, \ldots, m
\end{align*}
\]

- weak duality: \( p^* \geq d^* \) always
- strong duality: \( p^* = d^* \) for convex problem with constraint qualification (for example, Slater’s: primal problem is strictly feasible)
Semidefinite program

**Primal SDP** ($F_i, G \in S^k$)

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad x_1 F_1 + \cdots + x_n F_n \preceq G
\end{align*}
\]

- Lagrange multiplier is matrix $Z \in S^k$
- Lagrangian $L(x, Z) = c^T x + \text{Tr}(Z(x_1 F_1 + \cdots + x_n F_n - G))$
- Dual function

\[
g(Z) = \inf_x L(x, Z) = \begin{cases} 
- \text{Tr}(GZ) & \text{Tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\
-\infty & \text{otherwise}
\end{cases}
\]

**Dual SDP**

\[
\begin{align*}
\text{maximize} & \quad -\text{Tr}(GZ) \\
\text{subject to} & \quad Z \succeq 0, \quad \text{Tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n
\end{align*}
\]

$p^\star = d^\star$ if primal SDP is strictly feasible ($\exists x$ with $x_1 F_1 + \cdots + x_n F_n < G$)
Let's consider the following Second Order Cone Program (SOCP):

\[
\begin{align*}
\text{minimize} \quad & f^T x \\
\text{subject to} \quad & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m,
\end{align*}
\]

with variable \( x \in \mathbb{R}^n \). Let's show that the dual can be expressed as

\[
\begin{align*}
\text{maximize} \quad & \sum_{i=1}^{m} (b_i^T u_i + d_i v_i) \\
\text{subject to} \quad & \sum_{i=1}^{m} (A_i^T u_i + c_i v_i) + f = 0 \\
& \|u_i\|_2 \leq v_i, \quad i = 1, \ldots, m,
\end{align*}
\]

with variables \( u_i \in \mathbb{R}^{n_i}, v_i \in \mathbb{R}, \ i = 1, \ldots, m \) and problem data given by \( f \in \mathbb{R}^n, A_i \in \mathbb{R}^{n_i \times n}, b_i \in \mathbb{R}^{n_i}, c_i \in \mathbb{R} \) and \( d_i \in \mathbb{R} \).
We can derive the dual in the following two ways:

1. Introduce new variables $y_i \in \mathbb{R}^{n_i}$ and $t_i \in \mathbb{R}$ and equalities $y_i = A_i x + b_i$, $t_i = c_i^T x + d_i$, and derive the Lagrange dual.

2. Start from the conic formulation of the SOCP and use the conic dual. Use the fact that the second-order cone is self-dual:

\[ t \geq \|x\| \iff tv + x^T y \geq 0, \text{ for all } v, y \text{ such that } v \geq \|y\| \]

The condition $x^T y \leq tv$ is a simple Cauchy-Schwarz inequality.
We introduce new variables, and write the problem as

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad \|y_i\|_2 \leq t_i, \quad i = 1, \ldots, m \\
& \quad y_i = A_i x + b_i, \quad t_i = c_i^T x + d_i, \quad i = 1, \ldots, m
\end{align*}
\]

The Lagrangian is

\[
L(x, y, t, \lambda, \nu, \mu) = c^T x + \sum_{i=1}^{m} \lambda_i (\|y_i\|_2 - t_i) + \sum_{i=1}^{m} \nu_i^T (y_i - A_i x - b_i) + \sum_{i=1}^{m} \mu_i (t_i - c_i^T x - d_i)
\]

\[
= (c - \sum_{i=1}^{m} A_i^T \nu_i - \sum_{i=1}^{m} \mu_i c_i)^T x + \sum_{i=1}^{m} (\lambda_i \|y_i\|_2 + \nu_i^T y_i) + \sum_{i=1}^{m} (-\lambda_i + \mu_i) t_i
\]

\[-\sum_{i=1}^{n} (b_i^T \nu_i + d_i \mu_i).
\]
The minimum over $x$ is bounded below if and only if

$$\sum_{i=1}^{m} (A_i^T \nu_i + \mu_i c_i) = c.$$ 

To minimize over $y_i$, we note that

$$\inf_{y_i} (\lambda_i \| y_i \|_2 + \nu_i^T y_i) = \begin{cases} 0 & \| \nu_i \|_2 \leq \lambda_i \\ -\infty & \text{otherwise.} \end{cases}$$

The minimum over $t_i$ is bounded below if and only if $\lambda_i = \mu_i$. 
The Lagrange dual function is

\[
g(\lambda, \nu, \mu) = \begin{cases} 
- \sum_{i=1}^{n} (b_i^T \nu_i + d_i \mu_i) & \text{if } \sum_{i=1}^{m} (A_i^T \nu_i + \mu_i c_i) = c, \\
\|\nu_i\|_2 \leq \lambda_i, & \mu = \lambda \\
-\infty & \text{otherwise}
\end{cases}
\]

which leads to the dual problem

 maximize \(- \sum_{i=1}^{n} (b_i^T \nu_i + d_i \lambda_i)\)

 subject to \(\sum_{i=1}^{m} (A_i^T \nu_i + \lambda_i c_i) = c\)

 \(\|\nu_i\|_2 \leq \lambda_i, \quad i = 1, \ldots, m.\)

which is again an SOCP
Duality: SOCP

We can also express the SOCP as a **conic form** problem

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad -(c_i^T x + d_i, A_i x + b_i) \preceq_{K_i} 0, \quad i = 1, \ldots, m.
\end{align*}
\]

The Lagrangian is given by:

\[
L(x, u_i, v_i) = c^T x - \sum_i (A_i x + b_i)^T u_i - \sum_i (c_i^T x + d_i) v_i
\]

\[
= (c - \sum_i (A_i^T u_i + c_i v_i))^T x - \sum_i (b_i^T u_i + d_i v_i)
\]

for \((v_i, u_i) \succeq_{K_i} 0 \) (which is also \(v_i \geq \|u_i\|\))
With

\[ L(x, u_i, v_i) = \left( c - \sum_i (A_i^T u_i + c_i v_i) \right)^T x - \sum_i (b_i^T u_i + d_i v_i) \]

the dual function is given by:

\[ g(\lambda, \nu, \mu) = \begin{cases} 
- \sum_{i=1}^n (b_i^T \nu_i + d_i \mu_i) & \text{if } \sum_{i=1}^m (A_i^T \nu_i + \mu_i c_i) = c, \\
-\infty & \text{otherwise}
\end{cases} \]

The conic dual is then:

\[
\begin{align*}
\text{maximize} & \quad -\sum_{i=1}^n (b_i^T u_i + d_i v_i) \\
\text{subject to} & \quad \sum_{i=1}^m (A_i^T u_i + v_i c_i) = c \\
& \quad (v_i, u_i) \succeq K_i^*, \quad i = 1, \ldots, m.
\end{align*}
\]
Convex problem & constraint qualification

\[\Downarrow\]

Strong duality
Slater’s constraint qualification

Convex problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

The problem satisfies Slater’s condition if it is strictly feasible, i.e.,

\[\exists x \in \text{int} \mathcal{D} : \quad f_i(x) < 0, \quad i = 1, \ldots, m, \quad Ax = b\]

- also guarantees that the dual optimum is attained (if \( p^* > -\infty \))
- there exist many other types of constraint qualifications
KKT conditions for convex problem

If $\tilde{x}$, $\tilde{\lambda}$, $\tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal:

- from complementary slackness: $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$

hence, $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$ with $(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ feasible.

If Slater’s condition is satisfied, $x$ is optimal if and only if there exist $\lambda$, $\nu$ that satisfy KKT conditions

- Slater implies strong duality (more on this now), and dual optimum is attained
- generalizes optimality condition $\nabla f_0(x) = 0$ for unconstrained problem

Summary

- For a convex problem satisfying constraint qualification, the KKT conditions are necessary & sufficient conditions for optimality.
Proof

To simplify the analysis. We make two additional technical assumptions:

- The domain $\mathcal{D}$ has nonempty interior (hence, $\text{relint} \mathcal{D} = \text{int} \mathcal{D}$)
- We also assume that $A$ has full rank, i.e. $\text{Rank} A = p$. 
Proof

We define the set \( \mathcal{A} \) as

\[
\mathcal{A} = \{(u, v, t) \mid \exists x \in \mathcal{D}, \, f_i(x) \leq u_i, \, i = 1, \ldots, m, \, h_i(x) = v_i, \, i = 1, \ldots, p, \, f_0(x) \leq t\},
\]

which is the set of values taken by the constraint and objective functions.

If the problem is convex, \( \mathcal{A} \) is defined by a list of convex constraints hence is convex.

We define a second convex set \( \mathcal{B} \) as

\[
\mathcal{B} = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \mid s < p^*\}.
\]

The sets \( \mathcal{A} \) and \( \mathcal{B} \) do not intersect (otherwise \( p^* \) could not be optimal value of the problem).

First step: The hyperplane separating \( \mathcal{A} \) and \( \mathcal{B} \) defines a supporting hyperplane to \( \mathcal{A} \) at \( (0, p^*) \).
Illustration of strong duality proof, for a convex problem that satisfies Slater’s constraint qualification. The two sets \( \mathcal{A} \) and \( \mathcal{B} \) are convex and do not intersect, so they can be separated by a hyperplane. Slater’s constraint qualification guarantees that any separating hyperplane must be nonvertical.
Proof

By the separating hyperplane theorem there exists $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$ and $\alpha$ such that

\[(u, v, t) \in A \implies \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \geq \alpha, \quad (1)\]

and

\[(u, v, t) \in B \implies \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \leq \alpha. \quad (2)\]

From (1) we conclude that $\tilde{\lambda} \succeq 0$ and $\mu \geq 0$. (Otherwise $\tilde{\lambda}^T u + \mu t$ is unbounded below over $A$, contradicting (1).)

The condition (2) simply means that $\mu t \leq \alpha$ for all $t < p^*$, and hence, $\mu p^* \leq \alpha$.

Together with (1) we conclude that for any $x \in D$,

\[\mu p^* \leq \alpha \leq \mu f_0(x) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T (Ax - b) \quad (3)\]
Proof

Let us assume that $\mu > 0$ (separating hyperplane is nonvertical)

- We can divide the previous equation by $\mu$ to get

$$L(x, \tilde{\lambda}/\mu, \tilde{\nu}/\mu) \geq p^*$$

for all $x \in \mathcal{D}$

- Minimizing this inequality over $x$ produces $p^* \leq g(\lambda, \nu)$, where

$$\lambda = \tilde{\lambda}/\mu, \quad \nu = \tilde{\nu}/\mu.$$  

- By weak duality we have $g(\lambda, \nu) \leq p^*$, so in fact $g(\lambda, \nu) = p^*$.

This shows that strong duality holds, and that the dual optimum is attained, whenever $\mu > 0$. The normal vector has the form $(\lambda^*, 1)$ and produces the Lagrange multipliers.
Proof

Second step: Slater’s constraint qualification is used to establish that the hyperplane must be nonvertical, i.e. \( \mu > 0 \).

By contradiction, assume that \( \mu = 0 \). From (3), we conclude that for all \( x \in \mathcal{D} \),

\[
\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T (Ax - b) \geq 0.
\] (4)

- Applying this to the point \( \tilde{x} \) that satisfies the Slater condition, we have

\[
\sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) \geq 0.
\]

- Since \( f_i(\tilde{x}) < 0 \) and \( \tilde{\lambda}_i \geq 0 \), we conclude that \( \tilde{\lambda} = 0 \).
This is where we use the two technical assumptions.

- Then (4) implies that for all $x \in D$, $\tilde{v}^T(Ax - b) \geq 0$.
- But $\tilde{x}$ satisfies $\tilde{v}^T(A\tilde{x} - b) = 0$, and since $\tilde{x} \in \text{int } D$, there are points in $D$ with $\tilde{v}^T(Ax - b) < 0$ unless $A^T\tilde{v} = 0$.
- This contradicts our assumption that $\text{Rank } A = p$.

This means that we cannot have $\mu = 0$ and ends the proof.